AD-757 094

SEQUENTIAL ESTIMATION OF THE LARGEST NORMAL MEAN WHEN THE VARIANCE IS KNOWN

Saul Blumenthal

Cornell University

Prepared for:

Office of Naval Research
National Science Foundation

February 1973

AD757094

DEPARTMENT
OF
OPERATIONS RESEARCH

COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14850

R

DEPARTMENT OF OPERATIONS RESEARCH
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 171

February 1973

SEQUENTIAL ESTIMATION OF THE LARGEST NORMAL MEAN WHEN THE VARIANCE IS KNOWN

by

Saul Blumenthal*

*Cornell University and New York University

Supported in part by Office of Naval Research Contract ONR N00014-67-A-0077-0020 at the College of Engineering, Cornell University, and by National Science Foundation Grant GK 14073 at the School of Engineering and Science, New York University.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Cornell University Operations Research Department Ithaca, New York 14850	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

SEQUENTIAL ESTIMATION OF THE LARGEST NORMAL MEAN WHEN THE VARIANCE IS KNOWN

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Saul Blumenthal

6. REPORT DATE February, 1973	7a. TOTAL NO. OF PAGES ~~42~~ 46	7b. NO. OF REFS 3
8a. CONTRACT OR GRANT NO. N00014-67-A-0077-0020 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 171 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Mathematics Program Office of Naval Research Arlington, Virginia 22217

13. ABSTRACT

Given n observations from each of k normal populations with common known variance σ^2, the value of the largest of the k means (whose values are not known) is to be estimated using the largest value of the k sample means. It is desired to design a sampling rule which guarantees that the Mean Squared Error (M.S.E.) of the estimate does not exceed a given bound regardless of the configuration of values of the k means.

The M.S.E. is a function of $\bar{\Delta} = (\Delta_1, \ldots, \Delta_{k-1})$ where $\Delta_i = \max_{1 \le j \le k} \theta_j - \theta_i$ (Δ_k is zero by convention). If the Δ's are known a smaller sample size (fixed sample size procedures) can be used than the sample size n^* needed to guarantee the M.S.E. requirement for all $\bar{\Delta}$. Sequential and multi-sample procedures are considered which attempt to use sample information about $\bar{\Delta}$ to reduce sample sizes. Some properties of the sample size function and M.S.E. of these procedures are developed. Generally it is found that sample information about the value of $\bar{\Delta}$ is difficult to use efficiently.

DD FORM 1473 (PAGE 1)
1 NOV 65

S/N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
mathematical statistics						
ranking and selection						
sequential estimation						
ordered parameters						
multi-sample estimation						

1. INTRODUCTION AND SUMMARY

Given k normal populations with means $\theta_1,\ldots,\theta_k$ and common variance σ^2. Based on n observations from each population, a point estimate of $\theta^* = \max(\theta_1,\ldots,\theta_k)$ is desired. A natural estimator to use would be $X^* = \max(\bar{X}_1,\ldots,\bar{X}_k)$, where $\bar{X}_i$ $(1 \le i \le k)$ is the average of the n observations from the ith population. The risk or mean squared error function $R(X^*; \theta_1,\ldots,\theta_k)$ is easily seen to be

$$(1.1) \quad R(X^*; \theta_1,\ldots,\theta_k) = E[(X^*-\theta^*)^2|\theta_1,\ldots,\theta_k] = \frac{\sigma^2}{n}\int_{-\infty}^{\infty} y^2 d\Pi\Phi(y + (\sqrt{n}\Delta_i/\sigma))$$

where $\Delta_i = \theta^* - \theta_i$. It is to be noted that even if σ^2 is known, the risk depends on the unknown nuisance parameters $(\Delta_1,\ldots,\Delta_k)$ (one of which is zero). If a single-stage experiment is to be designed (n-chosen) so that $R(X^*; \theta_1,\ldots,\theta_k) \le r$ (where r is given), then an upper bound, say A^*, must be used for the integral in (1.1), and a conservative n must be chosen by equating r to $(\sigma^2 A^*/n)$. On the other hand, if multiple stage or sequential procedures can be used, then it might be hoped to reduce the total sample size by using information contained in the earlier stage samples regarding the values of the Δ_i. It is procedures of this type which will be investigated in this paper.

Since it is easy to see that $(X^*-\bar{X}_i)$ $(1 \le i \le k)$ is a collection of strongly consistent estimates of the Δ_i (see Dudewicz [2]), it would be hoped that when the desired sample size (obtained by equating (1.1) to r) is large, almost the same sample size would be obtained by replacing each Δ_i in (1.1) by its estimate. Certainly one would hope that the ratio of the two sample sizes would converge almost surely to unity. As will be seen, this

turns out to be the case if the desired sample size is made large by allowing r to go to zero, but not if σ^2 is allowed to become infinite while r is fixed. It will be seen in the latter case that the ratio mentioned has no almost sure limit. It does have a limiting distribution, and this will be examined. Most of the results on limiting distributions for the random sample size, and for the resulting risk of the estimator X^* based on random sample size have been obtained only for the case of two populations. Numerical evaluations of the limiting expected sample size and mean square error functions indicate that they do take some advantage of the savings possible when the Δ's are known, and may be of practical interest. Some simplified two sample procedures are examined and found not to be very good.

Estimation procedures for unknown σ^2 are considered in [1] in which a more extensive bibliography is given.

2. k POPULATIONS, CONVERGENCE RESULTS

It is easily seen that the supremum of (1.1) over $(\Delta_1,\dots,\Delta_k)$ is $(\sigma^2/n)A^*$ where

(2.1) $$A^* = A_k^* = \int_{-\infty}^{\infty} y^2 d\Phi^k(y)$$

(see Blumenthal [1], for example). The conservative approach to choosing n so that (1.1) is at most r (given), would be to take $n = n^*$ where

(2.2) $$n^* = A^*\sigma^2/r .$$

Clearly, if the Δ_i were known, a smaller sample size could be used by equating (1.1) to r. In this section we consider how much could be saved if the Δ_i

were known, and we examine multi-sample and sequential sampling procedures which try to accomplish these savings by using estimates of the Δ_i in their stead. Define for a given set $(\Delta_1,\ldots,\Delta_k)$

$$\eta^2 = (rn/A^*\sigma^2) \tag{2.3}$$

so that $(1/\eta^2)$ represents the "efficiency" possible if the Δ's are known, i.e., the ratio of the maximum sample size to the actual size needed for a given set of Δ's. Also, define (assuming that the one Δ_i having value zero is Δ_k)

$$x_i = \Delta_i n\sqrt{A^*}/\sqrt{r} \quad 1 \le i \le k-1, \quad x_k = 0 \tag{2.4}$$

and

$$H(x_1,\ldots,x_{k-1}) = (1/A^*) \int_{-\infty}^{\infty} y^2 d \prod_{i=1}^{k} \Phi(y+x_i) \ . \tag{2.5}$$

Then the result of equating (1.1) to r can be expressed as

$$\eta^2 = H(x_1,\ldots,x_{k-1}) \ . \tag{2.6}$$

Some of the properties of $H(x_1,\ldots,x_k)$ are examined in the Appendix. From Theorem 4.2.5 of Dudewicz [2], it follows that

$$R^* \overset{\text{(def.)}}{\equiv} \inf_{x_1,\ldots,x_{k-1}} H(x_1,\ldots,x_{k-1}) \geq (1/A^*)((1/2) + \int_{-\infty}^{0} y^2 d\Phi^k(y)) \geq (1/2\ A^*). \tag{2.7}$$

From (2.6) and (2.7), it is seen that $\inf \eta^2 = R^*$, or $\inf n = R^* n^*$. Thus while it is clear that savings are possible when the Δ_i are known, the percentage saving is limited by the fact that R^* is bounded from below.

A reasonable approach to taking advantage of the (limited) savings available would be after taking n observations to compute estimates $\hat{\Delta}_{i,n}$ of the Δ_i, then use these in (2.4), (2.5) and (2.6) to determine η^2, call this estimate $\hat{\eta}_n^2$, compute $\hat{n}$ as $\lceil n^*\hat{\eta}_n^2\rceil$ where $\lceil x\rceil$ is the smallest integer not less than x and take additional observations if $\hat{n} > n$, and stop sampling otherwise. A purely sequential procedure would repeat this operation after each stage of sampling, a multiple stage procedure having taken n_j observations from each population up to stage j and computing $\hat{n}_j$ based on these would observe at the next stage: $\max(0,\hat{n}-n_j)$ additional values from each population. In either case, it can be seen that it is always true that $\hat{n} > n_0$, where

$$n_0 = (R^*A^*\sigma^2/r) > (\sigma^2/2r) \ . \tag{2.8}$$

Thus for a sequential procedure, an initial sample size of n_0 must be taken before applying the stopping rule. For a multi-stage procedure it would be inefficient to use an initial sample less than n_0 thereby wasting the information available in the n_0 observations which must be taken. Therefore we assume a first sample of n_0 for any multi-stage rule. Define the sample size at stopping as N.

Let ρ_n^2 be defined as

$$\rho_n^2 = (n/n^*) \tag{2.9}$$

so that ρ_N^2 represents the ratio of the sample size at stopping to the maximum sample size (ignoring the possibility that n^* is not an integer). For two stage sampling where $N = \lceil n^*\hat{\eta}_{n_0}^2\rceil$, it is clear that

$$0 \le \rho_N^2 - \hat{\eta}_{n_0}^2 \le (1/n^*) \ . \tag{2.10}$$

For sequential and multi-stage procedures, it is possible to stop with $N \gg \hat{\eta}_N$, so that $\rho_N^2 \gg \hat{\eta}_N^2$ (the last computed value of $\hat{\eta}_n^2$). For multi-stage procedures, it is clear that

$$\rho_N^2 \geq \hat{\eta}_{n_0}^2 \tag{2.11}$$

because of the method of choosing the sample sizes.

Note that an alternative way to define a sequential procedure is to compute an estimate $\hat{R}_n$ of the risk at stage n based on substituting $\hat{\Delta}_{i,n}$ for Δ_i in (1.1), and to stop sampling the first time that $\hat{R}_n \leq r$. If R is a decreasing function of n for fixed $(\Delta_1,\ldots,\Delta_k)$, then the two approaches are equivalent. It would be surprising if R were not decreasing in n, but a proof exists only for $k = 2$ (see the Appendix). The alternative approach to multi-stage procedures is to compute $\hat{\Delta}_{i,n_j}$ based on the observations obtained through the jth stage, use this $\hat{\Delta}$ in place of Δ in (1.1), and solve for the n, say $\hat{n}$ which makes (1.1) equal r. At stage $(j+1)$, $(\hat{n}-n_j)$ additional observations are taken if $\hat{n} > n_j$, otherwise sampling stops.

The definition of N used in this paper assures that if $\hat{\Delta}_{i,n} \equiv \Delta_i$ for all n, then the resulting n will identically equal the result of equating (1.1) to r with the correct Δ_i in (1.1). This will be true regardless of whether R is a monotone decreasing function of n.

It is reasonable to ask whether ρ_N^2 has a stochastic limit if the sample size is forced to be large by making the desired risk r small. The answer is given by the following

Theorem 1. Given a sampling procedure with an initial sample size of n_0. Given strongly consistent estimates $\hat{\Delta}_{i,n}$ of the Δ_i. Let all Δ_i $(1 \leq i \leq k-1)$ be > 0. Then

(2.12) $$\lim_{r\to 0} \rho_N^2 = \lim_{r\to 0} \hat{\eta}_N^2 = (1/A^*) \quad \text{a.s.}$$

(2.13) $$\lim_{r\to 0} E(\rho_N^2) = \lim_{r\to 0} E(\hat{\eta}_N^2) = (1/A^*) \ .$$

Proof: By the strong consistency of the $\hat{\Delta}_{i,n}$, for given $\varepsilon > 0$, there is an M such that

(2.14) $$P\{\hat{\Delta}_{i,n} > (\Delta_i/2), \quad 1 \leq i \leq k, \quad \text{all} \quad n > M\} > 1-\varepsilon \ .$$

From (2.8), it is seen that for $r \leq r_0$ (say), $n_0 \geq M$ and thus defining $\hat{x}_{i,n}$ as $(\hat{\Delta}_{i,n} n\sqrt{A^*}/\sqrt{r})$ r_1 can be chosen so that for any arbitrarily large T (recalling that $n \geq \sqrt{R^*}$)

(2.15) $$P\{\hat{x}_{i,n} > T, \quad 1 \leq i \leq k, \quad \text{all} \quad n \geq n_0\} > 1-\varepsilon, \quad \text{if} \quad r < r_1 \ .$$

Since $\lim H(x_1,\ldots,x_{k-1})$ exists as all the x_i increase, and is $(1/A^*)$, from (2.15) and the fact that T is arbitrary, we conclude that for arbitrary $\delta > 0$,

(2.16) $$P\{(1/A^*)-\delta < \hat{\eta}_n^2 < (1/A^*), \quad \text{all} \quad n \geq n_0\} > 1-\varepsilon, \quad \text{if} \quad r < r_1 \ .$$

This completes the proof for $\hat{\eta}_N^2$.

For two stage procedures, (2.10) suffices to complete the proof for ρ_N^2. For multi-stage and sequential procedures, $\{\hat{\eta}_n^2 \geq (1/A^*)-\delta, \ n \geq n_0\}$ implies (with (2.11) and the preceding discussion) that $\{\rho_N^2 \geq (1/A^*)-\delta\}$. For multi-stage procedures, if stopping occurs at stage j, then $N = \hat{n}_{j-1}$, and $\rho_N^2 = \lceil n^*\hat{\eta}_{n_{j-1}}^2 \rceil / n^* < (1/A^*) + (1/n^*)$ if $\{\hat{\eta}_n^2 \leq (1/A^*), \ n \geq n_0\}$. This suffices

to give the result for the multi-stage case. For the sequential case, since stopping occurs at the first $n \geq \lceil n^* \hat{\rho}_n^2 \rceil$ so that $\{\hat{\rho}_n^2 \leq (1/A^*),\ n \geq n_0\}$, it follows that $\rho_N^2 \leq (1/A^*) + 2/n^*$. This completes the proof of (2.12).

The proof of (2.13) follows from the dominated convergence theorem.

If the $\hat{\Delta}_i$ are taken as $(X^* - \bar{X}_i)$ as mentioned in Section 1, it can be seen that (2.12) holds as $\sigma \to \infty$ provided that $\Delta_i = \Delta_i(\sigma)$ and $\lim_{\sigma\to\infty} \Delta_i(\sigma) = \infty$.

Further, note that Theorem 1 gives a very strong limitation on the usefulness of the proposed sequential procedure. When the Δ_i (> 0) are fixed, this procedure takes the same (order of magnitude) number of observations as would be the case if the Δ_i were all infinite. It should be noted, however, that even if the Δ_i were known, as $r \to 0$, ρ^2 goes to $(1/A^*)$. Thus the sequential (or multiple sample) rule does as well as can be done knowing the Δ's. If $k = 2$, $A^* = 1$, and this is no better than the conservative procedure. For $k \geq 3$, some savings are possible relative to the conservative procedure since $A^* > 1$. The situation when $\Delta_1 = \cdots = \Delta_{k-1} = 0$ is somewhat different as is shown below, in Theorem 2.

An additional point to remark here is that N for the procedures being considered can never exceed the conservative sample size $(A^*\sigma^2/r)$ (i.e., $\rho_N^2 \leq 1$). Thus, N has both lower and upper limits.

In examining (2.4), (2.5) and (2.6) it is seen that although a sizeable savings in sample size can be achieved for any value of r, the particular Δ_i values at which these savings are achievable depend on r. In particular, suppose $(x_1,\ldots,x_{k-1}) = \bar{x}$ is fixed. Then $\rho_{\bar{x}}^2$ can be computed from (2.6) and the Δ_i values at which this relative sample size will occur are given by

$$\Delta_i = \sqrt{r}\, x_i / \sqrt{A^* H(x_1,\ldots,x_k)} \quad . \tag{2.17}$$

If one fixes the value of $\bar{x}$ and allows the Δ_i to be given by (2.17), the question is whether the potential savings now available regardless of r can be achieved in any sense by the procedures which determine the sample size by using the estimates $\hat{\Delta}_{i,n}$ in place of the Δ_i. The following theorem gives a negative answer in the sense that ρ_N^2 does not converge with probability one to $\eta_{\bar{x}}^2$. In fact ρ_N^2 does not have an almost sure limit.

Consider the natural estimates of Δ_i, namely $(\bar{X}^*-\bar{X}_i)$. Note that $(\sqrt{n}/\sigma)(\bar{X}^*-\bar{X}_i)$ is distributed as (Y^*-Y_i) where the Y_i's are normal with means $(-\sqrt{n}\Delta_i/\sigma)$, and unit variance. If $n = c(A^*\sigma^2/r)$ (for any $c > 0$) and Δ_i is given by (2.17), these means are $(\sqrt{c}x_i/\eta_x)$, independently of r. For n of this form, the distribution of $(\sqrt{n}/\sigma)\hat{\Delta}_{i,n}$ (given that $\hat{\Delta}_{i,n} > 0$, i.e., that we are not looking at the one value of $\hat{\Delta}_{i,n}$ which by definition is zero) will assign positive weight to all intervals on the positive axis, and the weight will be independent of r. Further, the conditional joint distribution of the $(k-1)$ positive values of $(\sqrt{n}/\sigma)\hat{\Delta}_{i,n}$ (given that they are positive) assigns positive weight to all measurable $(k-1)$ dimensional sets having positive Lebesgue measure, and the weight is independent of r. Below we restrict attention to estimates $\hat{\Delta}_{i,n}$ having this property, which we refer to as property P.

<u>Theorem 2</u>. Let $\bar{x}$ be fixed and the Δ_i be given by (2.17). Let the estimates $\hat{\Delta}_{i,n}$ have property P defined in the preceding paragraph. Let $\eta_{\bar{x}}^2$ be given by (2.6) and $\hat{\eta}_N^2$ be as in Theorem 1, and ρ_N^2 be given by (2.9). Then for any $\bar{x}$ such that $R^* < \eta_{\bar{x}}^2 < 1$, there is an $\varepsilon > 0$ such that

$$\lim_{r\to 0} P\{|(\rho_N^2/\eta_{\bar{x}}^2)-1| \geq \varepsilon\} > 0 . \tag{2.18}$$

Proof: Let $\delta < 1$ be sufficiently large so that $H(x_1,\ldots,x_{k-1})$ is a decreasing function of each of its arguments on the set

$$S_\delta = \{(x_1,\ldots,x_{k-1}):\ \delta \le H(x_1,\ldots,x_{k-1}) \le 1;\ \ 0 \le x_i \le x_i^-(\delta),\ \ i = 1,\ldots,k-1\}$$

where $x_i^-(\delta)$ is the smaller of the two solutions to

$$\inf_{\{x_j,1\le j\le k-1,j\ne i\}} H(x_1,\ldots,x_{k-1}) = \delta$$

(if there are two solutions, i.e., if $\delta < (1/A^*)$). Letting $\Delta_i(\delta) = (x_i^-(\delta)/\sqrt{\delta})$, it can be seen from the monotonicity of $x^-(\delta)/\sqrt{\delta}$ (see Remark 2, Appendix) that the set of $(\Delta_1,\ldots,\Delta_{k-1})$ such that $\{\delta \le \eta^2 \le 1\}$ contains the set $T_\delta = (\sqrt{r/A^*}\,S_\delta/\sqrt{\delta})$ (i.e., each point in S_δ is multiplied by the given factor). Let $n = (cA^*\sigma^2/r)$ $(c > 0)$ and the Δ's be given by (2.17). Let $\bar{\Delta}_n = (\hat{\Delta}_{1,n},\ldots,\hat{\Delta}_{k-1,n})$ (assume $\hat{\Delta}_{k,n} = 0$ for notational convenience)

$$\{\hat{\eta}_n^2 \ge \delta\} \supset \{(\sqrt{n}/\sigma)\bar{\Delta} \in (\sqrt{n}/\sigma)T_\delta\} = \{(\sqrt{n}/\sigma)\bar{\Delta} \in \sqrt{c}\,S_\delta/\sqrt{\delta}\}\ .$$

From property P and the fact that $\sqrt{c}\,S_\delta/\sqrt{\delta}$ is measurable with positive Lebesgue measure, it follows that

$$P\{\hat{\eta}_n^2 \ge \delta\} > 0\ . \tag{2.19}$$

If n is taken to be n_0, then $c = R^*$ and using (2.11) along with (2.19) we see that for multi-stage procedures (two or more stages),

$$P\{\rho_N^2 \ge \delta\} > P\{\hat{\eta}_{n_0}^2 \ge \delta\} > 0\ . \tag{2.20}$$

If $\bar{x}$ is given, δ can be taken as $(1+\varepsilon)\eta_{\bar{x}}^2$, where ε is sufficiently large to satisfy the restriction on δ. This proves (2.18) for multi-stage procedures.

Consider now the sequential procedure. At stage n, it stops if $n \geq \hat{n}_n$, and since it can stop prior to reaching stage n,

$$P[N \leq n] > P[n \geq \hat{n}_n] \ . \tag{2.21}$$

Let $n = \lceil n^*\delta \rceil$. Now $\hat{n}_n = \lceil n^*\hat{\eta}_n^2 \rceil$. Noting that $\{\hat{\eta}_n^2 < \delta\} \Rightarrow \{n \geq \hat{n}_n\}$, and that $\{N \leq \lceil n^*\delta \rceil\} \Rightarrow \{\rho_N^2 \leq \delta + (1/n^*)\}$, we can conclude using (2.21) that

$$P\{\rho_N^2 \leq \delta+(1/n^*)\} > P\{\hat{\eta}_n^2 < \delta\}, \quad \text{when} \quad n = \lceil n^*\delta \rceil \ . \tag{2.22}$$

In this case let $\delta > R^*$ and sufficiently small so that we can choose the set S_δ such that it is connected, and on S_δ, $R^* \leq H(x_1,\ldots,x_{k-1}) \leq \delta$ and $H(x_1,\ldots,x_{k-1})$ is decreasing in all of its arguments. Clearly S_δ has positive Lebesgue measure and because of its definition, the set $(\Delta_1,\ldots,\Delta_{k-1})$ such that $\{R^* \leq \eta^2 \leq \delta\}$ contains $T_\delta = (\sqrt{r/A^*}\,U_\delta)$ and U_δ is obtained from S_δ by taking each $(x_1,\ldots,x_{k-1})$ on the contour $H(x_1,\ldots,x_{k-1}) = \eta^2$, $(R^* \leq \eta^2 \leq \delta)$ and transforming it to $(x_1/\eta,\ldots,x_{k-1}/\eta)$. The monotonicity of this transform on S_δ follows from Remark 2 in the Appendix. As before, with $c = \delta$, it follows that

$$P\{\hat{\eta}_n^2 \leq \delta\} \geq P\{(\sqrt{n}/\sigma)\bar{\Delta} \in \sqrt{\delta}\,U_\delta\} > 0, \quad \text{when} \quad n = \delta n^* \ . \tag{2.23}$$

Letting $\delta = (1-\varepsilon)\eta_{\bar{x}}^2$ where $\varepsilon > 0$ is chosen to satisfy the above requirements, (2.22) and (2.23) suffice to demonstrate (2.17) for the sequential case, completing the proof of the theorem.

<u>Remarks</u>:

1) The proof of Theorem 2 suggests that although N and ρ_N do not have a probability limit, they should have a distribution which may have a limit as $r \to 0$. For sequential and multi-stage procedures, the distribution would be very difficult to obtain analytically, but for two stage procedures it should be relatively simple. In view of (2.10) the limiting distribution of ρ_N^2 (as $n^* \to \infty$) will be the same as the distribution of $\hat{\eta}_{n_0}^2$ which is independent of n^*. Some aspects of this distribution will be studied for $k = 2$ in the next section.

2) Setting $\bar{x} = (0,\ldots,0)$ shows that the result of Theorem 1 does not hold when all Δ_i's are zero, and by remark (1) above, it can be expected that ρ_N^2 has a limiting distribution in this case. In fact, when any subset of p of the Δ's are zero while the remainder are fixed, the resulting ρ_N^2 will have a limit distribution found by studying the distribution of $H(\hat{x}_1,\ldots,\hat{x}_p,\infty,\ldots,\infty)$.

3) Large sample sizes are needed either if r is small as above or if σ^2 is large. The behavior of N and ρ_N^2 in the latter case can be studied by taking limits as $\sigma^2 \to \infty$ (σ^2 known). In this case for fixed $\Delta = (\Delta_1,\ldots,\Delta_{k-1})$, and fixed r it is clear that η^2 ($= \eta_\Delta^2$) is given as the solution of (2.4) and (2.5) and n is obtainable using (2.3). The value of (n/σ^2) is thus independent of σ^2. For given $\hat{\Delta}_{i,n}$ $(1 \le i \le k)$, it is also clear that $(\hat{n}/\sigma^2)$ has a value which is independent of σ^2 and which depends truly on the value of $\hat{\Delta}_{i,n}$. In this case it is very reasonable to ask whether ρ_N^2 converges to η_Δ^2 when the $\hat{\Delta}_{i,n}$ converge almost surely to the Δ_i (for fixed σ^2, as n increases).

For the rules under consideration, sample sizes are of the form $n = c\sigma^2$ $(c > 0)$ so that the nonzero values of $\{(\sqrt{n}/\sigma)\hat{\Delta}_{i,n},\ 1 \leq i \leq k\}$ have a conditional joint distribution independent of σ^2 which puts positive mass on any measurable set having positive Lebesgue measure. The method of proof of Theorem 2 gives as an immediate corollary, the following

Corollary 1. Let $\Delta = (\Delta_1,\ldots,\Delta_{k-1})$ be fixed. Let the distributional property P hold when $n = c\sigma^2$. Then for any Δ, such that $R^* < \eta_\Delta^2 < 1$, there is an $\varepsilon > 0$ such that

$$\lim_{\sigma^2\to\infty} P\{|(\rho_N^2/\eta_\Delta^2)-1| \geq \varepsilon\} > 0 . \tag{2.24}$$

It is clear in fact that $\hat{\eta}_{n_0}^2$ will have a distribution which does not depend on σ^2, and for the two sample case the limiting process is needed only to assure sufficiently large n so that the discrete variable ρ_N^2 will be close to the continuous $\hat{\eta}_N^2$.

It should be noted that the problem considered in this section differs in an important way from the case in which σ^2 is unknown, and the Δ_i are eliminated from consideration by some means (e.g., see Blumenthal [1]), and from the usual situations in which sequential or multi-stage estimation procedures are used. In these other cases, the desired sample size is an unbounded function of the unknown nuisance parameter, and consequently two-stage procedures tend to be inefficient relative to sequential procedures due to the possibility that the initial sample size may be far too small. In the problem considered here, the desired sample size is a bounded function of the nuisance parameter $(\Delta_1,\ldots,\Delta_{k-1})$, and in fact is bounded below as well as above. Thus it is possible to choose an initial sample size for a two stage procedure

such that no sequential procedure could stop with fewer observations, and such that this initial sample represents a sizeable fraction of the ultimate sample size (with a value of about 0.80 when $k = 2$ and a limit of $(1/2)$ as k increases to infinity). Thus the information available to determine the value of N is almost as great for the two sample procedure as for a sequential one, and little would appear to be gained by using a complex sequential stopping rule in place of the simple two sample one. This conclusion is supported by Theorems 1 and 2 which show both procedures to behave similarly in the limit.

3. TWO POPULATIONS; DISTRIBUTIONAL RESULTS FOR TWO STAGE SAMPLING

In this section, the distribution of the sample size and risk function will be studied for the two stage sampling procedure when either σ^2 is large and the Δ_i are fixed, or r is small and the Δ_i are proportional to $\sqrt{r}$. The discussion at the end of section 2 is taken as the rationale for not attempting the extremely difficult task of studying these for sequential or multi-stage procedures. Theorem 2 and Corollary 1 provide the incentive to discover just what the sample size and risk behavior is in these cases for which we do not have stochastic convergence of any sort.

Although it should be possible to characterize the risk and sample size distribution for any k, the analysis is greatly simplified when $k = 2$ (especially for the risk function) and we shall now specialize to the two population case. Some specialized notation will help in this study.

Let

(3.1) $$Z_{0,n} = (X_{1,n}+X_{2,n})/2; \quad Z_{1,n} = (X_{1,n}-X_{2,n})/2; \quad Z_n = |Z_{1,n}|$$

(3.2) $$\alpha = (\theta_1+\theta_2)/2; \quad \nu = (\theta_1-\theta_2)/2; \quad \omega = |\nu| .$$

Then

$$X_n^* = Z_{0,n} + Z_n, \quad \theta^* = \alpha + \omega \tag{3.3}$$

and $Z_{0,n}$ is normal, mean α, variance $(\sigma^2/2n)$ while $Z_{1,n}$ is normal mean ν, variance $(\sigma^2/2n)$ and $Z_{0,n}$ is independent of $Z_{1,n}$, hence of Z_n. Note that ω corresponds to $(\Delta_1/2)$ in the general notation and Z_n to $(\hat{\Delta}_{1,n}/2)$.

The risk (1.1) specializes to

$$(\sigma^2/n)(1 + 2F(x)) \tag{3.4}$$

where

$$F(x) = x^2\Phi(-x) - x\phi(x) \tag{3.5}$$

and

$$x = \omega\sqrt{2n}/\sigma = \eta\omega\sqrt{(2/r)} \tag{3.6}$$

with

$$\eta^2 = rn/\sigma^2 \quad \text{(since } A^* = 1) . \tag{3.7}$$

The function $F(x)$ is zero at $x = 0$, decreases to (-0.1012) at $x^* \cong 0.6120$, where x^* is the unique solution of $\phi(x) = 2x\Phi(-x)$, then rises again to zero, so that

$$R^* = \inf(1 + 2F(x)) = 1 - x^*\phi(x^*) = 0.7976 . \tag{3.8}$$

Corresponding to (2.6), we have

$$\eta^2 = (1 + 2F(x)) . \tag{3.9}$$

Based on an initial sample of $n_0 = R^*n^*$ observations (with $n^* = \sigma^2/r$), $\hat{\eta}^2$ is found from

(3.10) $$\hat{\eta}^2 = 1 + 2F(\hat{\eta} Z_{n_0} \sqrt{2/r}) \ .$$

Then, $N = \lceil n^*\hat{\eta}^2 \rceil$, and an additional $(N-n_0)$ observations are taken. The distribution of $\hat{\eta}^2$ is easily found. Re-write (3.6) as

(3.11) $$x = \eta\omega\sqrt{2n_0}/\sigma\sqrt{R^*}$$

and

(3.12) $$\hat{x} = \eta Z_{n_0} \sqrt{2n_0}/\sigma\sqrt{R^*} \ .$$

Further, for $R^* < \delta < 1$

(3.13) $$\{\hat{\eta}^2 < \delta\} \Leftrightarrow \{(x^-(\delta)/\sqrt{\delta})\sqrt{R^*} \leq Z_{n_0}\sqrt{2n_0}/\sigma \leq (x^+(\delta)/\sqrt{\delta})\sqrt{R^*}\}$$

where $x^-(\delta) < x^+(\delta)$ are the two solutions of $\delta = 1 + 2F(x)$. Note that $Z_{n_0}\sqrt{2n_0}/\sigma$ is $|W|$ where W is Normal, mean $(\sqrt{2n_0}/\sigma)\omega$, and variance 1. Henceforth, assume that

(3.14) $$(\sqrt{2n_0}/\sigma)\omega = \beta \ .$$

A convenient parametrization is obtained by letting

(3.15) $$\beta = x\sqrt{R^*/(1 + 2F(x))} \ .$$

This choice of β gives ω such that the solution of $\eta^2 = 1 + 2F(\eta\omega\sqrt{2n_0}/\sigma\sqrt{R^*})$ is simply, $\eta^2 = 1 + 2F(x)$. Note that this choice makes ω proportional to $\sqrt{r}$, so that if limits are taken as $r \to 0$, then ω is varying with r as

in Theorem 2, and if r is fixed and $\sigma^2 \to \infty$, then ω is fixed as in Corollary 1. Combining the above, when (3.14) holds,

$$P\{\hat{\eta}^2 < \delta\} = \Phi[(x^+\sqrt{R^*}/\sqrt{\delta})-\beta] - \Phi[(x^-\sqrt{R^*}/\sqrt{\delta})-\beta] \tag{3.16}$$
$$+ \Phi[-(x^-\sqrt{R^*}/\sqrt{\delta})-\beta] - \Phi[-(x^+\sqrt{R^*}/\sqrt{\delta})-\beta] \; .$$

From (2.10), we see that $\lim_{n^*\to\infty} P\{\rho_N^2 < \delta\}$ is given by (3.16). It is easily seen from the bounded convergence theorem and (2.10) that all moments (positive and negative) of ρ_N^2 converge to the corresponding moments of $\hat{\eta}^2$, as $n^* \to \infty$.

It is not especially convenient to use the distribution (3.16) to find the moments of $\hat{\eta}^2$ since it involves inverting $(1 + 2F(x))$ to obtain x^- and x^+. Let Y_1 be a standard normal variable. We can write $(\sqrt{2n_0}\, Z_{n_0}/\sigma)$ as $|Y_1 + \beta|$ (β given in (3.14)), and then $\hat{\eta}^2 = \eta^2(|Y_1 + \beta|)$ where $\eta^2(x)$ is the solution of

$$\eta^2 = 1 + 2F(\eta x/\sqrt{R^*}) \; . \tag{3.17}$$

Then

$$E\{(\hat{\eta}^2)^p\} = \int_{-\infty}^{\infty} (\eta^2|y+\beta|)^p \phi(y)dy = \int_{-\infty}^{-\beta} (\eta^2(-y-\beta))^p\phi(y)dy + \int_{-\beta}^{\infty} (\eta^2(y+\beta))^p\phi(y)dy \tag{3.18}$$
$$= \int_0^{\infty} (\eta^2(u))^p[\phi(u+\beta) + \phi(u-\beta)]du \; .$$

This integral still involves the implicitly defined function $\eta^2(u)$. Now, let

$$v = \eta u/\sqrt{R^*}, \quad \text{or} \quad u = \sqrt{R^*}v/\eta \tag{3.19}$$

so that

$$\eta^2 = 1 + 2F(v) \tag{3.20}$$

and

$$\frac{du}{dv} = (\sqrt{R^*}/\eta^2)(\eta - \frac{vF'(v)}{\eta}) = (\sqrt{R^*}/\eta^3)(1 + 2F(v) - vF'(v))$$

(3.21)

$$= (\sqrt{R^*}/\eta^3)(1 - v\phi(v)) .$$

Using this change of variables, (3.18) becomes

$$E\{(\hat{\eta}^2)^p\} = \sqrt{R^*} \int_0^\infty \eta^{(2p-3)}(v)(1 - v\phi(v))[\phi(\sqrt{R^*}(v/\eta(v))+\beta)$$

(3.22)

$$+ \phi(\sqrt{R^*}(v/\eta(v))-\beta)]dv$$

where $\eta(v)$ is now given explicitly by (3.20). For notational convenience, define

$$\gamma(v) = +\sqrt{R^*/(1 + 2F(v))} . \tag{3.23}$$

Then, (3.22) becomes

$$E\{(\hat{\eta}^2)^p\} = (R^*)^{(p-1)}\int_0^\infty \gamma^{3-2p}(v)(1 - v\phi(v))[\phi(v\gamma(v)+\beta) + \phi(v\gamma(v)-\beta)]dv . \tag{3.24}$$

Next the moments of $(X_N^*-\theta^*)$ will be studied, from which the risk can be derived. Using (3.3),

$$(X_N^*-\theta^*)^p = [(Z_{0,N}-\alpha) + (Z_N-\omega)]^p = \sum_{i=0}^{p} \binom{p}{i}(Z_{0,N}-\alpha)^i(Z_N-\omega)^{p-i} . \tag{3.25}$$

Also, conditioning on $N = n$, and noting that the stopping rule is based on Z_{n_0} and that Z_{0,n_0} is independent of Z_{n_0}, hence independent of N, it it seen that

$$E\{[(Z_{0,n}-\alpha)^i(Z_n-\omega)^{p-i}]|N=n\} = (\sigma/\sqrt{2n})^i E(W^i)E[(Z_n-\omega)^{p-i}|N=n] \tag{3.26}$$

hence

(3.27) $$E\{(Z_{0,N}-\alpha)^{i}(Z_N-\omega)^{p-i}\} = E(W^i)(\sigma/\sqrt{2})^{i}E[(Z_n-\omega)^{p-i}\,N^{-i/2}]$$

where W is a standard normal variable. Next, note that $Z_{1,N}$ can be rewritten as

$$\frac{1}{N}[n_0 Z_{1,n_0} + \tilde{n}Z_{1,\tilde{n}}]$$

where $\tilde{n} = N-n_0$ and $\tilde{n}Z_{1,\tilde{n}}$ is the sum of $\tilde{n}$ normal random variables, with mean ω and variance $\sigma^2/2$, which are independent of Z_{1,n_0} and of one another. Hence, we shall write

(3.28) $$Z_{1,N} = (1/N)[(\sqrt{n_0}\,\sigma/\sqrt{2})Y_1 + (\sqrt{\tilde{n}}\,\sigma/\sqrt{2})Y_2 + N\omega]$$

where Y_1 and Y_2 are independent standard normal variables. Using (2.9) and (3.14), we obtain

(3.29) $$Z_{1,N} = (r/2R^*)^{1/2}\{(R^*/\rho_N^2)Y_1 + [(R^*/\rho_N^2)(1 - (R^*/\rho_N^2))]^{1/2}Y_2 + \beta\} .$$

Note that $\hat{n}^2$ is given in (3.17) as a function of Y_1. Using (3.29), (2.10) and the fact that $R^* \leq (R^*/\rho_N^2) \leq 1$, the dominated convergence theorem yields for any j,

(3.30) $$\lim_{n^*\to\infty} E\{(Z_N/\sqrt{r})^j\} = (2R^*)^{-j/2}E\{|(R^*/\hat{n}^2)Y_1 + [(R^*/\hat{n}^2)(1 - R^*/\hat{n}^2)]^{1/2}\,Y_2 + \beta|^j$$

Let $\gamma(Y_1) = \gamma(|Y_1 + \beta|)$ be given by (3.23), (i.e., $\gamma^2 = (R^*/\hat{n}^2)$), and consider

$$E\{|\gamma^2(Y_1)Y_1 + [\gamma^2(Y_1)(1 - \gamma^2(Y_1))]^{1/2} Y_2 + \beta|\}$$

$$= EE\{|\gamma^2(y)y + [\gamma^2(y)(1 - \gamma^2(y))]^{1/2} Y_2 + \beta| \,|\, Y_1 = y\}$$

(3.31)
$$= E\{\int_{-\infty}^{-(A/B)} (-A-Bu)\phi(u)du + \int_{-(A/B)}^{\infty} (A+Bu)\phi(u)du\}$$

$$= E\{A(1 - 2\Phi(-A/B)) + 2B\phi(A/B)\}$$

where

(3.32)
$$A = Y_1\gamma^2(Y_1) + \beta, \quad B = [\gamma^2(Y_1)(1 - \gamma^2(Y_1))]^{1/2} .$$

Also,

(3.33)
$$E\{|\gamma^2(Y_1)Y_1 + [\gamma^2(Y_1)(1 - \gamma^2(Y_1))]^{1/2} Y_2 + \beta|^2\}$$

$$= EE\{A^2 + 2ABY_2 + B^2Y_2^2|Y_1\} = E\{A^2 + B^2\} .$$

Combining (2.8), (2.9), (3.14), (3.25), (3.27), (3.30), (3.31), (3.32), and (3.33) we obtain

(3.34)
$$\lim_{n^*\to\infty}(1/r)E(X_N^*-\theta^*)^2 = (1/2R^*)E\{Y_1^2\gamma^4(Y_1) + [\gamma^2(Y_1)(1 - \gamma^2(Y_1))]$$

$$+ 4\beta[A\Phi(-A/B) - B\phi(A/B)] + \gamma^2(Y_1)\} .$$

Note that the last $\gamma^2(Y_1)$ term in (3.34) represents the contribution of $Z_{0,N}$ to the risk, while the rest represents the contribution of Z_N. To evaluate the expectations in (3.34), it is necessary first to break the integrals into the two components $Y_1 + \beta \leq (\geq) 0$ as in (3.18). Some simplification is then possible by partial integrations, e.g.,

$$\int_0^\infty (y-\beta)^2\gamma^4(y)\phi(y-\beta)dy = -\beta\gamma^4(0)\phi(\beta) + \int_0^\infty [\gamma^4(y) + 4(y-\beta)\gamma^3(y)\frac{d}{dy}\gamma(y)]\phi(y-\beta)dy .$$

In this way, the $\gamma^4(Y_1)$ term can be eliminated. Lastly, the change of variables (3.19) can be used to get all functions defined explicitly (when this is done, the $(d\gamma(y)/dy)$ factor is absorbed into the differential for the new variable).

Using (3.30) and (3.31) it is also possible to evaluate the bias of the estimator X_N^* which by (3.17) and (3.27) is just $E(Z_N-\omega)$. We summarize the results as follows:

(3.35a) $$\lim_{n^*\to\infty}\sqrt{\frac{1}{r}}\, E(X_N^*-\theta^*) = (\sqrt{2}/R^{*3/2})[I_4 - I_3]$$

(3.35b) $$\lim_{n^*\to\infty}\frac{1}{r}E(Z_{0N}-\alpha)^2 = (1/2R^{*2})I_2$$

(3.35c) $$\lim_{n^*\to\infty}\frac{1}{r}E(Z_N-\omega)^2 = (1/2R^{*2})[4I_1 + I_2 + 4\beta I_3]$$

(3.35d) $$\lim_{n^*\to\infty}\frac{1}{r}E(X_N^*-\theta^*)^2 = (1/R^{*2})[2I_1 + I_2 + 2\beta I_3]$$

(3.35e) $$\lim_{n^*\to\infty}(1/n^*)E(N) = I_5$$

where

(3.36a) $$I_1 = \int_0^\infty [\phi(v) - 2v\Phi(-v)]\gamma^6(v)[C(v)\phi(C(v)) + D(v)\phi(D(v))]dv$$

(3.36b) $$I_2 = \int_0^\infty \gamma^5(v)[1 - v\phi(v)][\phi(C(v)) + \phi(D(v))]dv$$

(3.36c) $$\begin{aligned} I_3 = \int_0^\infty \{ &(\gamma^2(v)C(v)+\beta)\Phi(-(\gamma^2(v)C(v)+\beta)/B(v))\phi(C(v)) \\ &-(\gamma^2(v)D(v)-\beta)\Phi((\gamma^2(v)D(v)-\beta)/B(v))\phi(D(v)) \\ &-2B(v)\phi(v\gamma^2(v)/B(v))\phi(\beta/\gamma(v))\}\gamma^3(v)[1 - v\phi(v)]dv \end{aligned}$$

$$I_4 = \int_0^\infty [\phi(v) - 2v\Phi(-v)]\gamma^4(v)[\phi(C(v)) - \phi(D(v))]dv \tag{3.36d}$$

$$I_5 = \int_0^\infty \gamma(v)[1 - v\phi(v)][\phi(C(v)) + \phi(D(v))]dv \tag{3.36e}$$

and

$$\begin{aligned} B^2(v) &= \gamma^2(v)(1 - \gamma^2(v)) \\ C(v) &= v\gamma(v) - \beta \\ D(v) &= v\gamma(v) + \beta, \end{aligned} \tag{3.37}$$

with $\gamma(v)$ given by (3.23).

One question left unanswered is the following. Theorem 1 gives the almost sure limit of ρ_N^2 as $r \to 0$ when the Δ_i are fixed and positive. How does $E(X_N^*-\theta^*)^2$ behave in that case? The preceding development can be used to answer the question for $k = 2$ and the two sample procedure. The development of (3.34) can be used to write the exact expression

$$\begin{aligned} (1/r)E(X_N^*-\theta^*)^2 = (1/2R^*)E\{Y_1^2\gamma_N^4(Y_1) &+ [\gamma_N^2(Y_1)(1 - \gamma_N^2(Y_1))] \\ &+ 4\beta[A\Phi(-A/B) - B\phi(A/B)] + \gamma_N^2(Y_1)\} \end{aligned} \tag{3.38}$$

where $\gamma_N^2(Y_1) = R^*/\rho_N^2$, and where γ_N^2 is used for γ^2 in A and B. The pointwise limit of $\gamma_N^2(Y_1)$ as $\beta \to \infty$ (recall $\beta = \sqrt{2R^*/r}\ \omega$, ω fixed) is R^* and dominated convergence allows immediately taking limits under the expectation for all terms except $4\beta[A\Phi(-A/B) - B\phi(A/B)]$. This term has a pointwise limit of zero as β increases, as is easily seen from the standard tail approximation to the normal distribution, and it is easy to verify that the expectation has a limit of zero. Thus,

$$\lim_{r\to 0} E(X_N^*-\theta^*)^2 = \lim_{\beta\to\infty} (3.35d) = 1 . \tag{3.39}$$

In general (k populations, sequential or multiple sampling), it is conjectured that the risk when the Δ_i are fixed and $r \to 0$ will be the same as if a non-random number $n = (n^*/A^*)$ of observations were taken. This would give

$$\lim_{r\to 0} E(1/r)(X_n^*-\theta^*)^2 = \lim_{r\to 0} \int_{-\infty}^{\infty} y^2 d\Pi\Phi(y + (\Delta_i/\sqrt{A^*r})) = 1 .$$

Since the integrals in (3.36) are quite complex, it is very difficult to tell from looking at them how the expected sample size and risk behave, except in the limiting case, where $\beta \to \infty$. In that case, $\eta^2(|Y + \beta|)$ approaches unity for all Y and dominated convergence allows the conclusion that $\lim_{\beta\to\infty} E\{\eta^2(|Y + \beta|)\} = \lim_{\beta\to\infty} E\{\eta^{-2}(|Y + \beta|)\} = 1$ so that (3.35b) approaches $(1/2)$ and (3.35e) approaches unity as β increases. We have seen already that (3.35d) approaches unity and the same argument shows that (3.35a) approaches zero and (3.35c) approaches $(1/2)$. For finite β, since $\eta^2(|Y + \beta|) < 1$, it can be concluded that (3.35b) will exceed $(1/2)$ and that (3.35e) will be less than unity. Otherwise, the behavior is not ascertainable from examination of the formulas. Therefore numerical integrations have been performed, and the results are given in the following section.

4. NUMERICAL RESULTS

The performance of the two sample procedure as given by (3.35) was evaluated numerically for a range of values of ω. If ω were known then n could be chosen so that (3.4) equals r (given) and the value n_ω defined in this way can be considered as an ideal value under perfect information. To

evaluate the effectiveness of the two sample procedure, $E(N)$ should be compared to n_ω, and $\frac{1}{r}E(X_N^*-\theta^*)^2$ should be compared to unity which is the value of $\frac{1}{r}E(X_{n_\omega}^*-\theta^*)^2$. Similarly the biases of the ideal and two sample procedures can be compared. A convenient reparametrization is the following. Let x be a running parameter and define

$$(4.1) \qquad rn_x = \sigma^2(1+2F(x)), \quad \text{and} \quad \omega_x = x/\sqrt{2n_x} \ .$$

Clearly n_x is n_{ω_x}. The operating characteristics of the two sample procedure have been computed for $\omega = \omega_x$, as x traverses a suitable range.

Table 1 gives for each x, λ_x where $\lambda_x = \omega_x\sqrt{n^*}$ and $n^* = \sigma^2/r$. This is followed by $n_x^0 = n_x/n^*$ and $\bar{N}_x = \lim\limits_{n^*\to\infty}(E(N)/n^*)$ (as given by (3.35)). Next is $\bar{M}_x = \lim\limits_{n^*\to\infty}\frac{1}{r}E(X_N^*-\theta^*)^2$, and its two constituent components $\bar{M}_{0x} = \lim\limits_{n^*\to\infty}\frac{1}{r}E(Z_{0N}-\alpha)^2$ and $\bar{M}_{1x} = \lim\limits_{n^*\to\infty}\frac{1}{r}E(Z_N-\omega)^2$. After this is $B_I = \sqrt{\frac{1}{r}}E(X_{n_x}^*-\theta^*)$, the normalized bias of the ideal or perfect information procedure ($B_I = -2\lambda_x F(x)/x^2$), and $\bar{B}_x = \lim\limits_{n^*\to\infty}\sqrt{\frac{1}{r}}E(X_N^*-\theta^*)$. Finally for reference, the normalized mean squared error (M_c) and bias (B_c) of the conservative single sample procedure whose sample size is n^* are given (evaluated at $\omega = \omega_x$, so that $M_c = 1+2F(\sqrt{2}\,\lambda_x)$, and $B_c = -F(\sqrt{2}\,\lambda_x)/\lambda_x$).

Comparison of $E(N)$ with n_x shows that $E(N)$ tends to be flatter as a function of x, lying below n_x at extreme values and above in the central region. The M.S.E. of X_N^* is not quite as flat as might have been hoped. Note however that it is below unity at $x = 0.20$ and at $x = 1.50$ even though $E(N) < n_x$, indicating that the two sample procedure may be somewhat more effective in using the observations taken than is the one sample procedure. Consider the M.S.E. of a single sample procedure whose sample size $\bar{n}$ equals $E(N)$ of

Table 1

Comparative Mean Squared Errors and Biases

x	$\hat{\ }_x$	M^{0}_x	$\bar{N}_x$	$\bar{M}_x$	$\bar{M}_{0x}$	$\bar{M}_{1x}$	B_I	$\bar{B}_x$	M_c	B_c
.00	.0000	1.0000	.8662	1.1028	.5801	.5227	.5642	.5327	1.0000	.5642
.05	.0360	.9626	.8662	1.0667	.5801	.4866	.5397	.4950	.9619	.5289
.10	.0733	.9298	.8664	1.0337	.5800	.4538	.5148	.4573	.9276	.4939
.15	.1117	.9015	.8667	1.0044	.5798	.4245	.4892	.4201	.8973	.4595
.20	.1510	.8772	.8671	.9786	.5795	.3991	.4634	.3836	.8714	.4260
.25	.1910	.8568	.8676	.9567	.5792	.3775	.4375	.3481	.8496	.3937
.30	.2315	.8399	.8683	.9383	.5788	.3596	.4116	.3139	.8321	.3627
.35	.2723	.8263	.8692	.9236	.5782	.3453	.3860	.2813	.8185	.3332
.40	.3132	.8156	.8702	.9121	.5776	.3345	.3608	.2503	.8087	.3055
.45	.3541	.8077	.8713	.9037	.5769	.3267	.3362	.2212	.8021	.2794
.50	.3947	.8022	.8725	.8980	.5762	.3218	.3123	.1940	.7986	.2551
.55	.4351	.7989	.8738	.8947	.5753	.3194	.2892	.1687	.7975	.2326
.60	.4751	.7976	.8753	.8936	.5744	.3191	.2671	.1454	.7987	.2119
.65	.5145	.7980	.8768	.8942	.5735	.3207	.2460	.1239	.8016	.1928
.70	.5534	.8000	.8785	.8963	.5725	.3239	.2259	.1043	.8060	.1753
.75	.5917	.8032	.8802	.8997	.5714	.3283	.2070	.0865	.8115	.1592
.80	.6294	.8077	.8820	.9041	.5703	.3338	.1892	.0703	.8180	.1446
.85	.6666	.8130	.8838	.9092	.5691	.3401	.1725	.0557	.8250	.1313
.90	.7031	.8192	.8858	.9150	.5679	.3471	.1569	.0425	.8325	.1191
.95	.7391	.8260	.8877	.9213	.5667	.3546	.1425	.0306	.8403	.1080
1.00	.7746	.8334	.8897	.9279	.5655	.3624	.1291	.0200	.8483	.0979
1.05	.8096	.8411	.8918	.9347	.5642	.3705	.1167	.0105	.8563	.0887
1.10	.8441	.8490	.8939	.9417	.5629	.3788	.1053	.0021	.8643	.0804
1.15	.8783	.8572	.8960	.9488	.5616	.3872	.0949	-.0054	.8722	.0727
1.20	.9122	.8654	.8982	.9559	.5603	.3956	.0853	-.0120	.8800	.0658
1.25	.9457	.8735	.9004	.9629	.5589	.4040	.0765	-.0177	.8876	.0594
1.30	.9790	.8816	.9026	.9699	.5576	.4123	.0686	-.0228	.8950	.0536
1.35	1.0121	.8896	.9049	.9768	.5562	.4206	.0613	-.0271	.9021	.0484
1.40	1.0451	.8973	.9072	.9835	.5548	.4287	.0547	-.0308	.9089	.0436
1.45	1.0779	.9048	.9095	.9900	.5534	.4366	.0488	-.0339	.9155	.0392
1.50	1.1106	.9121	.9118	.9963	.5520	.4444	.0434	-.0365	.9218	.0352
1.55	1.1433	.9190	.9141	1.0025	.5505	.4519	.0385	-.0386	.9278	.0316
1.60	1.1759	.9256	.9165	1.0084	.5491	.4592	.0342	-.0402	.9335	.0283
1.65	1.2086	.9319	.9189	1.0140	.5477	.4663	.0302	-.0415	.9389	.0253
1.70	1.2413	.9378	.9213	1.0194	.5462	.4731	.0267	-.0423	.9440	.0226
1.75	1.2740	.9434	.9237	1.0245	.5448	.4797	.0235	-.0428	.9488	.0201
1.80	1.3068	.9486	.9261	1.0293	.5433	.4860	.0207	-.0430	.9533	.0179
1.85	1.3397	.9535	.9285	1.0338	.5419	.4919	.0182	-.0429	.9575	.0159
1.90	1.3726	.9580	.9309	1.0380	.5404	.4976	.0160	-.0425	.9615	.0140
1.95	1.4057	.9622	.9333	1.0419	.5390	.5029	.0140	-.0419	.9651	.0124
2.00	1.4389	.9660	.9357	1.0455	.5375	.5080	.0122	-.0411	.9685	.0109
2.50	1.7767	.9900	.9583	1.0643	.5240	.5403	.0028	-.0269	.9903	.0027
3.00	2.1238	.9977	.9764	1.0570	.5134	.5436	.0005	-.0120	.9977	.0005
3.50	2.4754	.9996	.9883	1.0398	.5065	.5332	.0001	-.0037	.9996	.0001
4.00	2.8285	.9999	.9949	1.0224	.5028	.5196	.0000	-.0005	.9999	.0000

the two sample procedure. Table 2 gives a few selected values.

Table 2

x	$(\bar{n}/n^*)$	$n^*E(X^*_{\bar{n}} - \theta^*)^2$	$\bar{M}_x$
.00	.867	1.151	1.103
.15	.867	1.040	1.004
1.00	.890	0.942	0.928
1.55	.914	1.005	1.002
2.00	.936	1.025	1.046
2.50	.958	1.030	1.064
4.00	.995	1.005	1.022

It is clear that for small x, the two sample procedure uses its observations more effectively than does a comparable single sample one, but for larger x values the M.S.E. of X^*_N decreases more slowly toward its asymptote of unity and gives slightly higher M.S.E. values than the comparable single sample estimate.

A similar behavior in the bias of the two sample procedure relative to B_I is also noted.

It was demonstrated that (N/n_x) does not converge stochastically to unity as n^* increases, and the numerical results show that $(E(N)/n_x)$ is not too close to unity either. Table 1 also shows that $MSE(X^*_N)$ fails to achieve the goal of being constant at unity. In spite of these facts, it is seen that the two sample procedure does take some advantage of the possible savings available when ω is known, and the M.S.E. curve does not rise very far above unity. Compared to the conservative procedure, about a 10% saving in sample size is achievable for moderate values of x, and the two sample procedure may very well be acceptable in practice.

A simplified two sample procedure can be constructed in the following way. Take n_1 observations and compute Z_{n_1}. Divide $(0,\infty)$ into m regions, $R_1,\ldots,R_m$. If $Z_{n_1} \in R_j$, let $N = n_j$ $(1 \le j \le m)$, and take $\tilde{n} = N-n_1$ additional observations. (It may be desirable to take $n_1 > n_0$ for small m). We assume that the n_j are chosen as $\lceil n_1/\gamma_j \rceil$ where $\gamma_1 = 1,\ldots,\gamma_m = (n_1/n^*)(>0)$ are fixed constants. This makes $n_m = \lceil n^* \rceil$. The regions R_j will be chosen as follows: R_1 is the interval $[\sigma C_1^-/\sqrt{2n_1}, \ \sigma C_1^+/\sqrt{2n_1}]$, R_j $(2 \le j \le m)$ is the pair of intervals $[\sigma C_j^-/\sqrt{2n_1}, \ \sigma C_{j-1}^-/\sqrt{2n_1}) \cup (\sigma C_{j-1}^+/\sqrt{2n_1}, \ \sigma C_j^+/\sqrt{2n_1}]$ where

$$(4.2) \qquad C_i^- = x^-(\delta_i)\sqrt{\gamma_m/\delta_i}\ , \qquad C_i^+ = x^+(\delta_i)\sqrt{\gamma_m/\delta_i} \qquad (1 \le i \le m)$$

and the values $R^* \le \delta_1 \le \cdots \delta_{m-1} < \delta_m = 1$ $(x^-(1) = 0,\ x^+(1) = \infty)$ are given. This choice of R_j is such that if $\hat{n}$ is the solution of $r = (\sigma^2/\hat{n})(1 + 2F(Z_{n_1}\sqrt{2\hat{n}}/\sigma))$, then $\hat{n} \le \delta_i n^*$ for $(\sigma C_i^-/\sqrt{2n_1}) \le Z_{n_1} \le (\sigma C_i^+/\sqrt{2n_1})$. A conservative choice of sample sizes n_j would be given by $n_j > \hat{n}$ or $(\gamma_m/\gamma_j) > \delta_j$ $(1 \le j \le m)$. When $m = \lceil n^* \rceil - \lceil n_0 \rceil$ and $n_i = \lceil n_0 \rceil + (i-1)$, $\delta_i = (n_i/n^*)$, then this is equivalent to the previous two sample procedure. Define

$$(4.3) \qquad \gamma_N^2 = n_1/N\ .$$

As $n^* \to \infty$ (and $n_1 \to \infty$ since $\gamma_m > 0$), the range of possible values for γ_N^2 converges to $(\gamma_1,\ldots,\gamma_m)$. Clearly, as n^* increases, γ_N^2 converges in law to a random variable γ_∞^2 whose distribution is

$$(4.4) \qquad P\{\gamma_\infty^2 = \gamma_j\} = [\Phi(x+\beta) + \Phi(x-\beta)]\Big|_{I_j} \qquad 1 \le j \le m$$

where $F(x)|_{I_j}$ denotes

$$(4.5)\qquad \begin{cases} F(C_1^+) - F(C_1^-) & j = 1 \\ F(C_j^+) - F(C_{j-1}^+) + F(C_{j-1}^-) - F(C_j^-) & j = 2,\ldots,m \end{cases}$$

and where

$$(4.6)\qquad \beta = \sqrt{2n_1}\,\omega/\sigma\ .$$

Using (4.4), it is easy to see that

$$(4.7)\qquad \lim_{n^*\to\infty}(1/n^*)E(N) = \gamma_m \lim_{n^*\to\infty} E(1/\gamma_N^2) = \gamma_m E(1/\gamma_\infty^2),$$

and using (4.4) and (3.27) that

$$(4.8)\qquad \lim_{n^*\to\infty} \frac{1}{r}E(Z_{0N}-\alpha)^2 = (1/2)\lim_{n^*\to\infty} E(n^*/N) = (1/2\ \gamma_m)\lim_{n^*\to\infty} E(\gamma_N^2) = (1/2\ \gamma_m)E(\gamma_\infty^2)\ .$$

The other expressions corresponding to (3.35) and (3.36) simplify in a similar way to give

$$(4.9)\qquad \lim_{n^*\to\infty}\sqrt{\frac{1}{r}}\,E(X_N^*-\theta^*) = (1/\sqrt{\gamma_m})[I_8 - \sqrt{2}\,I_7]$$

$$(4.10)\qquad \lim_{n^*\to\infty}\frac{1}{r}E(Z_N-\omega)^2 = (1/2\ \gamma_m)[E(\gamma_\infty^2) + I_6 + 4\beta I_7]$$

$$(4.11)\qquad \lim_{n^*\to\infty}\frac{1}{r}E(X_N^*-\theta^*)^2 = (1/2\ \gamma_m)[2E(\gamma_\infty^2) + I_6 + 4\beta I_7]$$

where

$$I_6 = -\sum_{k=1}^{m}\gamma_k^2[(z-\beta)\phi(z-\beta) + (z+\beta)\phi(z+\beta)]\Big|_{I_k}$$

$$I_7 = \sum_{k=1}^{m} \{\gamma_k[\phi(z+\beta)\Phi((z-\beta\lambda_k)/\sqrt{\lambda_k}) - \phi(z-\beta)\Phi(-(z+\beta\lambda_k)/\sqrt{\lambda_k})]\Big|_{I_k}$$

$$-2\sqrt{\gamma_k}\ \phi(\beta/\sqrt{\gamma_k})\Phi\left(\frac{z}{\sqrt{1-\gamma_k}}\right)\Bigg|_{I_k}$$

$$+ \beta \int_{I_k} [\phi(y+\beta)\Phi((y-\beta\lambda_k)/\sqrt{\lambda_k}) + \phi(y-\beta)\Phi(-(y+\beta\lambda_k)/\sqrt{\lambda_k})]dy\}$$

$$I_8 = \sum_{k=1}^{m} \gamma_k[\phi(z+\beta) - \phi(z-\beta)]\Big|_{I_k}$$

and

$$\lambda_k = (1 - \gamma_k)/\gamma_k$$

$$\int_{I_k} f(x)dx = \int_{C_1^-}^{C_1^+} f(x)dx \quad \text{if} \quad k = 1$$

$$= \int_{C_k^-}^{C_{k-1}^-} f(x)dx + \int_{C_{k-1}^+}^{C_k^+} f(x)dx \quad \text{if} \quad k = 2,\dots,m\ .$$

In the particularly simple case in which $m = 2$, we denote γ_m by A, and C_1^- by C and C_1^+ by d, the formulas (4.7) through (4.11) reduce to

(4.12a) $$\lim_{n^*\to\infty}(1/n^*)E(N) = 1 - I_{10}$$

(4.12b) $$\lim_{n^*\to\infty}\sqrt{\frac{1}{r}}\,E(X_N^*-\theta^*) = (1/\sqrt{A})[I_{12} - \sqrt{2}\ I_{11}]$$

(4.12c) $$\lim_{n^*\to\infty}\frac{1}{r}E(Z_{0N}-\alpha)^2 = (1/2A)[A + I_{10}]$$

(4.12d) $$\lim_{n^*\to\infty}\frac{1}{r}E(Z_N-\omega)^2 = (1/2A)[A + I_{10} + 2I_9 + 4\beta I_{11}]$$

(4.12e) $$\lim_{n^*\to\infty} \frac{1}{r}E(X_N^* - \theta^*)^2 = (1/A)[A + I_{10} + I_9 + 2\beta I_{11}]$$

where,

$$I_9 = ((1-A^2)/2)[(c+\beta)\phi(c+\beta) - (d+\beta)\phi(d+\beta) + (c-\beta)\phi(c-\beta) - (d-\beta)\phi(d-\beta)]$$

$$I_{10} = (1-A)[\Phi(d-\beta) - \Phi(c-\beta) + \Phi(-c-\beta) - \Phi(-d-\beta)$$

$$I_{11} = \beta\{\Phi(-\beta/\sqrt{A}) - [(\int_{-d}^{-c} + \int_{c}^{d})\Phi\left(\frac{-(y+\beta\lambda)}{\sqrt{\lambda}}\right)\phi(y-\beta)dy]\}$$

$$-\sqrt{A}\ \phi(\beta/\sqrt{A})\{1 - 2[\Phi\left(\frac{d}{\sqrt{1-A}}\right) - \Phi\left(\frac{c}{\sqrt{1-A}}\right)]\}$$

$$-A\{\phi(d+\beta)\Phi((d-\beta\lambda)/\sqrt{\lambda}) + \phi(c-\beta)\Phi(-(c+\beta\lambda)/\sqrt{\lambda})$$

$$-\phi(c+\beta)\Phi((c-\beta\lambda)/\sqrt{\lambda}) - \phi(d-\beta)\Phi(-(d+\beta\lambda)/\sqrt{\lambda})\}$$

$$+ [\phi(d+\beta) - \phi(c+\beta)] + \beta[\Phi(d+\beta) - \Phi(c+\beta)]$$

$$I_{12} = ((1-A)/\sqrt{2})[\phi(d+\beta) + \phi(c-\beta) - \phi(d-\beta) - \phi(c+\beta)]$$

$$\lambda = (1-A)/A\ .$$

Numerical evaluations of (4.12) were made for a few combinations of A and δ $(= \delta_1)$ (on which C_1 is based through (4.2)), namely:

A \ δ			
0.85	0.85		
0.90	0.85	0.90	
0.95	0.85	0.90	0.95

The values of $x^-(\delta)$ and $x^+(\delta)$ needed for constructing the procedures are

δ	$x^-(\delta)$	$x^+(\delta)$
0.85	.26903	1.10597
0.90	.15282	1.41758
0.95	.06836	1.81395

Tables 3 and 4 show the values of $\bar{N}$ (given (4.12a)) and $\bar{M}$ (given by (4.12e)) respectively for the various combinations of A and δ shown above. For reference, $\bar{N}$ and $\bar{M}$ for the "regular" two sample procedure have been transferred from Table 1. These computations suggest that the simplified two sample procedures sacrifice too much of the information in the first sample resulting in very poor M.S.E. characteristics with little compensating saving in sample size. Generally, they seem inferior even to the conservative procedure, and do not pick up any of the good features of the regular two sample estimator.

5. ALTERNATE LOSS FUNCTION

Another possible approach to constructing a sequential procedure for the problem of Section 2, is to let the loss function for an estimate based on n observations be $cn + (X^*-\theta^*)^2$ where c represents a relative cost per observation. The risk is then $cn + (A^*\sigma^2/n)H(x_1,\ldots,x_{k-1})$ $(x_i = \Delta_i\sqrt{n}/\sigma, \quad 1 \le i \le k-1)$. The optimal sample size is obtained by differentiating the risk with respect to n and is the solution of

(5.1) $$c = (A^*\sigma^2/2n^2)G(x_1,\ldots,x_{k-1})$$

where $G(x_1,\ldots,x_{k-1})$ is given by (A.11). Define

Table 3

Comparison of $\bar{N}$ for Alternative 2-sample Procedures

x	Regular	A: .85 δ: .85	.90 .85	.90 .90	.95 .85	.95 .90	.95 .95
0	.8662	.9221	.9481	.9278	.9740	.9639	.9562
.05	.8662	.9222	.9481	.9278	.9741	.9639	.9562
.10	.8664	.9223	.9482	.9279	.9741	.9640	.9563
.20	.8671	.9229	.9486	.9284	.9743	.9642	.9565
.30	.8683	.9240	.9494	.9292	.9747	.9646	.9568
.40	.8702	.9255	.9504	.9304	.9753	.9653	.9573
.50	.8725	.9274	.9518	.9319	.9760	.9660	.9579
.75	.8802	.9337	.9562	.9369	.9783	.9687	.9601
1.00	.8897	.9412	.9615	.9431	.9811	.9719	.9630
1.25	.9004	.9492	.9670	.9499	.9839	.9755	.9665
1.50	.9118	.9573	.9726	.9571	.9868	.9792	.9703
1.75	.9237	.9653	.9780	.9644	.9896	.9830	.9744
2.00	.9357	.9728	.9831	.9715	.9921	.9866	.9787
3.00	.9764	.9935	.9963	.9926	.9984	.9968	.9934
4.00	.9949	.9993	.9997	.9991	.9999	.9997	.9991

Table 4

Comparison of $\bar{M}$ for Alternative 2-sample Procedures

x	Regular	A: .85 δ: .85	.90 .85	.90 .90	.95 .85	.95 .90	.95 .95
0	1.1028	1.0473	1.0290	1.0493	1.0134	1.0229	1.0346
.05	1.0667	1.0353	1.0140	1.0275	.9921	.9965	1.0041
.10	1.0337	1.0305	1.0062	1.0138	.9779	.9780	.9822
.20	.9786	1.0427	1.0129	1.0114	.9721	.9666	.9654
.30	.9383	1.0829	1.0484	1.0424	.9964	.9899	.9846
.40	.9121	1.1466	1.1087	1.1031	1.0476	1.0449	1.0364
.50	.8980	1.2274	1.1878	1.1875	1.1200	1.1260	1.1150
.75	.8997	1.4613	1.4235	1.4572	1.3506	1.3975	1.3848
1.00	.9279	1.6818	1.6507	1.7416	1.5837	1.6928	1.6939
1.25	.9629	1.8539	1.8290	1.9923	1.7714	1.9554	1.9956
1.50	.9963	1.9671	1.9447	2.1870	1.8945	2.1578	2.2664
1.75	1.0245	2.0178	1.9927	2.3119	1.9456	2.2832	2.4867
2.00	1.0455	2.0058	1.9733	2.3566	1.9248	2.3204	2.6345
3.00	1.0570	1.5255	1.4674	1.8266	1.4125	1.7461	2.2809
4.00	1.0224	1.1047	1.0802	1.1843	1.0611	1.1447	1.3598

(5.2) $$\nu^2 = (n^2 c/\sigma^2),$$

so that

(5.3) $$x_i = \Delta_i \sqrt{\nu}/\sqrt{\sigma}\ \sqrt[4]{c} \ .$$

A sequential stopping rule which is natural to use when the Δ_i are not known, is stop the first time that

(5.4) $$c \geq (A^*\sigma^2/2n^2)G(\hat{x}_1,\ldots,\hat{x}_{k-1})$$

where

$$\hat{x}_i = \hat{\Delta}_{i,n}\sqrt{n}/\sigma \qquad 1 \leq i \leq k\text{-}1 \ .$$

If one studies the behavior of such a rule for fixed σ^2 as $c \to 0$, it is seen from comparing (5.1), (5.2) and (5.3) with (2.3), (2.4) and (2.6) that the behavior is almost identical, with $\sqrt{c}$ replacing r and G in place of H. It might have been suspected that the difficulties pointed up by Theorems 1 and 2 were due to the choice of loss function and criterion for choosing the sample size. It might then have been hoped that using this loss function would lead to probability one convergence to unity (as c decreases) of the ratio of the random sample size N to the optimal sample size (computed from (5.1) with fixed Δ_i). Instead, the analogues of Theorems 1 and 2 hold as $c \to 0$ (in the analogue of Theorem 1, the sample size behaves as though all $\Delta_i = +\infty$, which from (A.12) implies that $\lim_{c\to 0}(N\sqrt{c}/\sigma) = 1$ a.s.). Suppose c is fixed and $\sigma \to \infty$, and it is assumed that $\{(\sqrt{n}/\sigma)\hat{\Delta}_{i,n},\ 1 \leq i \leq k\}$ have the joint distribution of $\{(Y^*-Y_i),\ 1 \leq i \leq k\}$ where the Y's are normal, variance 1 and means $\{-\sqrt{n}/\sigma)\Delta_i\}$. For the present sampling rule it is seen that $(\sqrt{n}/\sigma)$

is proportional to $(1/\sqrt{\sigma})$ so that the means of the Y_i are in the limit $(\sigma \to \infty)$ zero, regardless of the original Δ_i. Thus as $\sigma \to \infty$ for fixed Δ_i, N has a limiting distribution (not an almost sure limit) which is independent of the Δ_i. A result like Theorem 2 will obtain if the Δ_i are proportional to $\sqrt{\sigma}$.

If the global maximum of G is at the origin (see (A.15) and the discussion following it), then the conservative sampling procedure takes $n = (\sigma\sqrt{A^*}/\sqrt{c})$ (see (A.13)).

In the case $k = 2$, the shape of G is very similar to that of H and all of the results of section 3 could be transcribed easily for this loss function.

6. A RELATED PROBLEM

Let $X_1,\ldots,X_n$ be normal, mean ω, variance τ^2, and suppose that it is desired to estimate $|\omega|$ by means of the estimator $Z_n = |\bar{X}_n|$. The risk of this estimator is

$$R(Z,\omega) = (\tau^2/n)[1 + 4F(\omega\sqrt{n}/\tau)] \tag{6.1}$$

where $F(x)$ is given by (3.5). The behavior of this risk function is the same as of (3.4) so that sequential and multiple sample procedures based on substituting Z_n for ω in (6.1) will have the same properties as described in Sections 2 and 3. The bias, expected sample size, and risk functions for the two sample procedure will be very similar to those tabulated for the case $k = 2$, but not identical since for (6.1), $R^* = 0.596$, and from equations (3.36), it is seen that R^* enters these expressions in a non-linear way.

ACKNOWLEDGMENT

The author wishes to thank Mr. Daniel Camerini for programming and computing the numerical results in this paper.

REFERENCES

[1] Blumenthal, S. "Sequential Estimation of the Largest Normal Mean when the Variance is Unknown," Technical Report No. 164, Dept. of Operations Research, Cornell University, (1973).

[2] Dudewicz, E. "Estimation of Ordered Parameters," Technical Report No. 60, Dept. of Operations Research, Cornell University, (1969).

[3] Saxena, K.M.L. and Tong, Y. L. "Interval Estimation of the Largest Mean of k Normal Populations with Known Variances," J. Amer. Statist. Assoc. 64, No. 325 (1969), 296-299.

APPENDIX

PROPERTIES OF THE RISK FUNCTION

The risk function $R(X^*;\ \theta_1,\ldots,\theta_k)$ can be expressed as

$$\text{(A.1)}\quad E(X^*-\theta^*)^2 = \int_{-\infty}^{\infty} (x-\theta^*)^2 d \prod_{i=1}^{k} \Phi(\sqrt{n}\,(x-\theta_i)/\sigma)$$

$$= \int_{-\infty}^{\infty} y^2 d \prod_{i=1}^{k} \Phi(\sqrt{n}\,(y+\Delta_i)/\sigma)$$

$$= \int_{0}^{\infty} y^2 d\Big[\prod_{i=1}^{k} \Phi(\sqrt{n}(y+\Delta_i)/\sigma) - \prod_{i=1}^{k} \Phi(\sqrt{n}\,(-y+\Delta_i)/\sigma)\Big]$$

$$= 2\int_{0}^{\infty} y\Big[1 - \prod_{i=1}^{k} \Phi(\sqrt{n}\,(y+\Delta_i)/\sigma) + \prod_{i=1}^{k} \Phi(\sqrt{n}\,(-y+\Delta_i)/\sigma)\Big]dy$$

$$= (\sigma^2/n)2\int_{0}^{\infty} y\Big[1 - \prod_{i=1}^{k} \Phi(y+x_i) + \prod_{i=1}^{k} \Phi(-y+x_i)\Big]dy$$

$$= (A^*\sigma^2/n)H(x_1,\ldots,x_{k-1})$$

where

$$\text{(A.2)}\quad \Delta_i = \theta^* - \theta_i,\quad x_i = (\sqrt{n}\,\Delta_i/\sigma)\ .$$

<u>Lemma A.1</u>. For fixed values of $(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_{k-1})$ $H(x_1,\ldots,x_{k-1})$ taken as a function of x_i decreases for $0 \leq x_i < x_i^*$ then increases for $x_i^* < x_i < \infty$.

<u>Proof</u>: Writing

$$\text{(A.3)}\quad \Big(\frac{1}{A^*}\Big)\frac{\partial H(x_1,\ldots,x_k)}{\partial x_i} = 2e^{-x_i^2/2}\int_0^{\infty} y\phi(y)\Big[e^{x_i y}\prod_{j\neq i}\Phi(-y+x_j)-e^{-x_i y}\prod_{j\neq i}\Phi(y+x_j)\Big]dy$$

it is seen that at $x_i = 0$, the bracketed term in the integral is negative for all $y > 0$ so that the derivative is also. Further, a lower bound on the bracketed term is $[(1 + x_i y)\Phi^{k-1}(-y)-1]$ so that the integral in (A.3) becomes infinite as x_i increases. Taking the derivative of the integral in (A.3) with respect to x_i gives

$$\left(\frac{1}{A^*}\right)\frac{\partial}{\partial x_i}\left(\frac{1}{2}e^{x_i^2/2}\frac{\partial H(x_1,\ldots,x_k)}{\partial x_i}\right) = \int_0^\infty y^2\phi(y)[e^{x_i y}\prod_{j\neq i}\Phi(-y+x_j) + e^{-x_i y}\prod_{j\neq i}\Phi(y+x_j)]dy>0$$

so that the derivative has only one sign change, completing the proof.

Two of the more important properties of the risk function for $k = 2$ will now be demonstrated.

Lemma A.2.

$$\frac{\partial R}{\partial n}(X^*;\ \theta_1,\theta_2) < 0\ . \tag{A.4}$$

Proof: From (3.4), (3.5), and (3.6) it is seen that

$$\frac{\partial R}{\partial n} = -(\sigma^2/n^2)[1 + 2F(x) - 2nf(x)(x/2n)] \tag{A.5}$$

where $f(x) = dF(x)/dx$, and it is easily checked that (A.5) becomes

$$\frac{\partial R}{\partial n} = -(\sigma^2)/n^2)(1 - x\phi(x))$$

completing the proof since $x\phi(x) < \phi(1) < 1$ for $x \geq 0$.

In obtaining the distributions in section 3, use was made of the equivalence between the statements

$$\{\eta^2 < \delta\} \quad \text{where} \quad \eta^2 \quad \text{satisfies:} \quad \eta^2 = [1 + 2F(\eta\omega\sqrt{2/r})]$$

and $\{\omega_\delta^- < \omega < \omega_\delta^+\}$, where $\omega_\delta = x_\delta/\sqrt{\delta}$, and x_δ is one of the two solutions of

$$\delta = 1 + 2F(x) \ . \tag{A.6}$$

This equivalence depends on the fact that $\omega_\delta^-(\omega_\delta^+)$ is a monotone increasing (decreasing) function of δ $(0 < \delta < R^*)$. The fact that $x_\delta^-(x_\delta^+)$ is a monotone increasing (decreasing) function of δ follows immediately from the fact that $(1 + 2F(x))$ decreases monotonely for $x < x^*$ and increases monotonely for $x > x^*$ (see Lemma A.1). We now show that ω_δ behaves in the same way.

<u>Lemma A.3</u>. Let ω_δ^- and ω_δ^+ be the two solutions of

$$\delta = 1 + 2F(\omega\sqrt{2\delta/r}) \ . \tag{A.7}$$

Then $\frac{\partial\omega^{-(+)}}{\partial\delta}$ has the same sign as $\frac{\partial x_\delta^{-(+)}}{\partial\delta}$, and is zero only when the latter is zero.

<u>Proof</u>: Writing ω_δ as $c(x_\delta/\sqrt{\delta})$, it is seen that

$$\frac{\partial\omega_\delta}{\partial\delta} = \frac{c}{2\delta^{3/2}}(-x_\delta + 2\delta\,\frac{dx_\delta}{d\delta}) \ . \tag{A.8}$$

From (A.6) it is seen that

$$1 = 2f(x_\delta)\frac{dx_\delta}{d\delta} \tag{A.9}$$

Thus using (A.9) in (A.8) along with (A.6) it is seen that

$$\text{(A.10)}\quad \frac{\partial \omega_\delta}{\partial \delta} = \frac{c}{\delta^{3/2}(2f(x_\delta))}[-x_\delta f(x_\delta) + 1+2F(x_\delta)] = \frac{c}{\delta^{3/2}}(1 - x\psi(x))\left(\frac{\partial x_\delta}{\partial \delta}\right)$$

and since $(1 - x\phi(x)) > 0$, this completes the proof.

It is difficult to characterize $H(x_1,\ldots,x_{k-1})$ in great detail, but the following remarks indicate some of the known properties.

Remarks:

1) By direct application of the proof of Theorem 1 of Saxena and Tong [3], it is easy to see that for fixed $(x_1,\ldots,x_p)$ (say)

$$\max_{x_{p+1},\ldots,x_{k-1}} H(x_1,\ldots,x_{k-1}) = H(x_1,\ldots,x_p,0,\ldots,0)\ .$$

2) The unimodal shape of $\bar{H}(x_i) = H(x_1,\ldots,x_k)$ in x_i (fixed values of $(k-2)$ other x's) implies monotonicity of the solutions $x_i^-(\delta)$, $x_i^+(\delta)$ to the equation $\delta = \bar{H}(x_i)$ (for $\delta > 1/A^*$, there will be only one solution, $x_i^-(\delta)$). As in the case $k = 2$, it would be desirable to be able to show monotonicity of $(x_i^{-(+)}(\delta)/\sqrt{\delta})$, or alternatively to show that $x_i/\sqrt{\bar{H}(x_i)}$ is monotone increasing in x_i, i.e., to show that $[2\bar{H}(x_i) - x_i(d\bar{H}(x_i)/dx_i)] > 0$. This is trivial when $\bar{H}(x_i)$ is decreasing, but illusive in general. As x_i becomes very large, $\bar{H}(x_i)$ has a positive limit, while $x_i(d\bar{H}(x_i)/dx_i)$ goes to zero. Thus the difficulty encountered is for $x_i > x_i^*$, but not too large.

3) Similarly, it would be desirable to show that for general k, $(dR(X^*;\theta_1,\ldots,\theta_k)/dn) < 0$. This is equivalent to showing that $G(x_1,\ldots,x_{k-1}) > 0$

for all $(x_1,\ldots,x_{k-1})$ where

$$\text{(A.11)}\quad G(x_1,\ldots,x_{k-1}) = [2H(x_1,\ldots,x_k) - \sum_{i=1}^{k-1} x_i(\partial H(x_1,\ldots,x_k)/\partial x_i)] \; .$$

The difficulties are of the same nature as mentioned above.

4) It is not clear whether the function $H(x_1,\ldots,x_{k-1})$ has a unique minimum. The system of equations $\partial H(x_1,\ldots,x_k)/\partial x_i = 0$ $(1 \le i \le k-1)$ has a solution on the diagonal $x_1 = x_2 = \cdots = x_{k-1} = x$, given by

$$\int_0^\infty y\phi(y)[e^{xy}\phi^{k-1}(-y+x)-e^{-xy}\phi^{k-1}(y+x)]dy = 0 \; .$$

Whether this solution is unique and whether there are solutions off the diagonal is not clear. If a solution off the diagonal exists, then from the symmetry of the function, any permutation of that solution is also a solution. If one considers a ray $x_i = tr_i$ $(r_i \ge 0$, fixed, $0 \le t < \infty$, $1 \le i \le k-1)$

$$\frac{\partial H(tr_1,\ldots,tr_{k-1},0)}{\partial t} = \sum_{i=1}^{k-1} r_i\left(\frac{\partial H(x_1,\ldots,x_k)}{\partial x_i}\bigg|_{x_i=tr_i}\right) \; .$$

It is clear that this derivative is initially negative since each term is, and that there is a t^* (depending on the r's) such that for $t > t^*$, it is positive but since the individual terms change signs at different values of t and since the terms are not necessarily monotone increasing (because of the $e^{-(tr_i)^2/2}$ factor) it is not clear whether there is only one sign change. There is a possibility of the function oscillating in a limited t range.

5) One property which can be established is that R^* is a decreasing function of k, with a lower bound which approaches $(1/2A^*)$ as k increases. To see this, let $x_1^*,\ldots,x_{k-1}^*$ be such that

$$H(x_1^*,\ldots,x_{k-1}^*) = \min_{x_1,\ldots,x_{k-1}} H(x_1,\ldots,x_{k-1}) = R_k^* .$$

Then

$$R_{k+1}^* = \min_{x_1,\ldots,x_k} H(x_1,\ldots,x_k) \leq \min_{x_k} H(x_1^*,\ldots,x_{k-1}^*,x_k)$$

$$< H(x_1^*,\ldots,x_{k-1}^*,\infty) = H(x_1^*,\ldots,x_{k-1}^*) = R_k^* .$$

From Theorem 4.2.5 of Dudewicz [2], it is easily seen that

$$(1/A^*)[(1/2) + \int_{-\infty}^{0} x^2 d\Phi^k(x)] \leq R_k^* \leq (1/A^*)$$

and the lower bound decreases to $(1/2A^*)$ as k increases. A somewhat sharper upper bound for R_k^* is

$$(1/A^*) \inf_{0 \leq x < \infty} \int_{-\infty}^{\infty} y^2 d\Phi(y)\Phi^{k-1}(y+x) .$$

The function $G(x_1,\ldots,x_{k-1})$ of (A.11) is of interest in section 5, and a few of its properties will be described here. Clearly, since $(\partial H(x_1,\ldots,x_{k-1})/\partial x_i)$ goes exponentially fast to zero as x_i increases,

$$\text{(A.12)} \qquad \lim_{\min(x_1,\ldots,x_{k-1})\to\infty} G(x_1,\ldots,x_{k-1}) = 2H(\infty,\ldots,\infty) = (2/A^*)$$

and

$$G(0,\ldots,0) = 2H(0,\ldots,0) = 2 \geq (2/A^*) \quad \text{(A.13)}$$

Also

$$\frac{\partial G(x_1,\ldots,x_{k-1})}{\partial x_j} = \frac{\partial H(x_1,\ldots,x_{k-1})}{\partial x_j} - \sum_{i=1}^{k-1} x_i \frac{\partial^2 H(x_1,\ldots,x_{k-1})}{\partial x_i \partial x_j} \quad \text{(A.14)}$$

so that

$$\left.\frac{\partial G(x_1,\ldots,x_{k-1})}{\partial x_j}\right|_{(0,\ldots,0)} = \left.\frac{\partial H(x_1,\ldots,x_{k-1})}{\partial x_j}\right|_{(0,\ldots,0)} < 0 \quad \text{(A.15)}$$

Thus $G(x_1,\ldots,x_{k-1})$ decreases from its value at the origin which in turn is higher than the value at infinity. Whether the value at the origin is a universal maximum, and details on the shape of the function are not known. When $k = 2$, $G(x) = 2(1 - x\phi(x))$ decreases monotonely for $0 \leq x \leq 1$, then increases for $x \geq 1$ with a minimum value of $2(1 - \phi(1)) \approx 2(0.76)$.